



OPEN

Resource efficient aortic distensibility calculation by end to end spatiotemporal learning of aortic lumen from multicentre multivendor multidisease CMR images

Tuan Aqeel Bohoran¹, Kelly S. Parke², Matthew P. M. Graham-Brown², Mitul Meisuria², Anvesha Singh², Joanne Wormleighton², David Adlam², Deepa Gopalan³, Melanie J. Davies⁴, Bryan Williams⁵, Morris Brown⁶, Gerry P. McCann^{2,7} & Archontis Giannakidis^{1,7}✉

Aortic distensibility (AD) is important for the prognosis of multiple cardiovascular diseases. We propose a novel resource-efficient deep learning (DL) model, inspired by the bi-directional ConvLSTM U-Net with densely connected convolutions, to perform end-to-end hierarchical learning of the aorta from cine cardiovascular MRI towards streamlining AD quantification. Unlike current DL aortic segmentation approaches, our pipeline: (i) performs simultaneous spatio-temporal learning of the video input, (ii) combines the feature maps from the encoder and decoder using non-linear functions, and (iii) takes into account the high class imbalance. By using multi-centre multi-vendor data from a highly heterogeneous patient cohort, we demonstrate that the proposed method outperforms the state-of-the-art method in terms of accuracy and at the same time it consumes ~ 3.9 times less fuel and generates ~ 2.8 less carbon emissions. Our model could provide a valuable tool for exploring genome-wide associations of the AD with the cognitive performance in large-scale biomedical databases. By making energy usage and carbon emissions explicit, the presented work aligns with efforts to keep DL's energy requirements and carbon cost in check. The improved resource efficiency of our pipeline might open up the more systematic DL-powered evaluation of the MRI-derived aortic stiffness.

Clinical backdrop

The aorta is the main artery of the human body operating as a conduit that forwards oxygenated blood to peripheral organs and tissues. The elastic buffering capacity of the aortic wall is of vital importance to the cardiovascular system function as it caters to the transformation of blood flow from pulsatile (as generated by the left ventricular contraction) to steady (as required by the periphery)¹. However, a progressive loss of the aortic wall elasticity manifests itself naturally with age². Factors such as hypertension³, diabetes⁴, connective tissue disorders⁵, protein genetic variations⁶, and congenital heart defects^{7,8} may further escalate this process. The

¹School of Science and Technology, Nottingham Trent University, Nottingham NG11 8NS, UK. ²Department of Cardiovascular Sciences, University of Leicester and the NIHR Leicester Biomedical Research Centre, Glenfield Hospital, Leicester LE3 9QP, UK. ³Imperial College London & Cambridge University Hospitals, Cambridge CB2 0QQ, UK. ⁴Leicester Diabetes Centre, University of Leicester and the NIHR Leicester Biomedical Research Centre, Leicester General Hospital, Leicester LE5 4PW, UK. ⁵Institute of Cardiovascular Science, University College London (UCL), National Institute for Health Research (NIHR), UCL Hospitals Biomedical Research Centre, London WC1E 6DD, UK. ⁶Department of Clinical Pharmacology, William Harvey Research Institute, Queen Mary University of London, London EC1M 6BQ, UK. ⁷These authors jointly supervised this work: Gerry P. McCann and Archontis Giannakidis. ✉email: archontis.giannakidis@ntu.ac.uk

increased aortic stiffness has been shown to be an early marker of vascular aging and a powerful independent predictor of adverse cardiovascular events and mortality in various cohorts^{9–13}.

Aortic distensibility (AD) is a direct measure of aortic stiffness and is defined as the maximum relative change in the aortic lumen cross-sectional area (A) during the cardiac cycle for a given pressure step at fixed vessel length

$$AD \left(10^{-3} \text{ mmHg}^{-1}\right) = \frac{A_{max} - A_{min}}{A_{min} \times PP} \quad (1)$$

where PP is the pulse pressure (= systolic blood pressure – diastolic blood pressure), and A_{max} , A_{min} denote the maximum and minimum areas, respectively^{14,15}. AD is inversely proportional to the square of the pulse wave velocity³. Several reports^{12,16–18} have highlighted the value of AD as an indicator of aortic stiffness.

Cardiovascular magnetic resonance imaging (CMR) is recognised as the gold-standard non-invasive test for quantification of ventricular volumes and mass. Increasingly, CMR is used for the semantic segmentation-based calculation of AD from electrocardiogram-gated steady-state free precession (SSFP) cine images acquired in the plane perpendicular to the thoracic aorta at the level of the pulmonary artery bifurcation. This calculation is achieved by delineating the aortic lumen in the above 2D cine images, working out the minimum and maximum areas across the cardiac cycle by multiplying in each area the number of pixels by the pixel dimensions, and then combining those with a separately measured PP, as described in Eq. (1). Unlike other modalities used to assess aortic stiffness, CMR provides a high-resolution visualisation of the aorta in both spatial and temporal domains. Along with the reproducible placement of the imaging plane perpendicular to the vessel, CMR also permits the local stiffness evaluation at numerous aorta sections in the same study³. The CMR assessment of aortic stiffness has been validated against invasive intra-aortic pressure measurements¹⁹.

The challenge

However, the image processing-based workflows for calculating AD from aortic cine CMR images are suboptimal in current clinical practice^{20–22}, as they rely on semi-automated methods for the segmentation of the ascending aorta (AAo) and descending aorta (DAo) cross-sectional areas throughout the cardiac cycle. As well as being time-consuming, the image interpretation step of such pipelines is also subject to intra- and inter-observer variability. Additionally, the CMR expertise is costly. To enhance the clinical applicability of the CMR-derived AD and, thus, foster the efficient management of individuals with increased aortic wall stiffness, fast fully-automated methods, which also improve the robustness of the aortic lumen area quantification during the cardiac cycle, are needed. The automation to the cine CMR image interpretation analysis in the AD calculation is nonetheless challenging due to: (i) the high cross-sectional aortic shape variability across the cardiac cycle, different patients, and diverse pathologies, and (ii) the aorta brightness heterogeneity as a result of blood flow. In addition, the CMR acquisition protocols are very diverse among different studies and institutions which further adds to the task's complexity²³.

Related work

Two recent papers deployed deep learning (DL)-based approaches to fully automate the aortic lumen segmentation from cine CMR images^{24,25}. DL or hierarchical representations learning is a rapidly growing branch of machine learning in which the models learn complex raw-input-data representations that are tuned to the specific task at hand²⁶. The deployment of DL models has been a game changer in computer vision (among a wide range of other industries) as it has permitted to achieve or even surpass human-level performance in a number of visual tasks including image classification, object detection, and semantic segmentation²⁷. Recent research has highlighted the superb capabilities of DL models to analyse CMR images^{28,29}. However, the two aorta DL studies treated the task at hand as a sparse annotation problem and evaluated their method only on a very limited number of cardiac cycle time frames (namely the end diastole (ED) and end systole (ES)) for which the ground truth labels were available. The targets of the remaining time frames (that were also used to map the input data during training) could not be visually validated and were acquired by pipelines which are likely to suffer from registration errors or poor convergence of the active contour algorithm. On top of this, the previous studies either completely ignored the temporal continuity inherent in the cine image sequence by treating the image segmentation task as static²⁵, or simulated the use of time by stacking the recurrent part after the convolutional layers²⁴. However, correlated spatio-temporal features cannot be learnt when spatial and temporal features are explicitly determined in separate regions of the network³⁰. In addition, the previous papers combined the feature maps from the encoder and decoder using simple concatenation. Nonetheless, it has been argued that such a practice results in less precise segmentation than when non-linear functions are employed for this task³¹. Next, the loss function used by previous studies neglected the fact that the region of interest (ROI) class is significantly smaller than the background class. Moreover, the previous papers analysed datasets acquired by following a single data acquisition protocol on a relatively healthy cohort. Finally, previous work completely ignored the resource efficiency matter of the proposed pipelines. However, this is an important issue as the substantial computation and energy demands by DL models go together with a considerable environmental and financial cost³². Improving the efficiency of algorithms should be a high priority in DL research alongside accuracy³².

Overview of the proposed method

In this paper, we propose to enhance the AD calculation by performing aortic lumen segmentation throughout the cardiac cycle using of a novel resource-efficient spatio-temporal DL model, inspired by the bi-directional ConvLSTM (BConvLSTM) U-Net with densely connected convolutions³¹. Our network is trained from scratch and evaluated on an aortic cine CMR dataset that was fully annotated by experts. The present paper is the first

work to perform end-to-end (i.e., over the entire cardiac cycle) hierarchical learning and testing of the aortic lumen area from cine CMR images. Our approach joins the temporal with the spatial processing of the video input by merging the encoder and decoder feature maps through a BConvLSTM (non-linear) unit³³. We employ the focal Tversky loss during training which is better suited for problems with a high class imbalance in the data³⁴. We use multi-centre multi-vendor data from a highly heterogeneous patient cohort which significantly adds to the generalisation power of the proposed aortic lumen segmentation algorithm. Finally, the network we propose in this study is resource-efficient to help promote environmentally friendly and more inclusive DL research and practices. We perform quantitative evaluations of the energy consumption and carbon cost during training. It is worth noting that resource-efficient approaches are also more well-suited towards deployment in hardware-constrained platforms, which in turn will open up the more widespread DL-powered evaluation of the CMR-derived aortic stiffness. To examine the impact brought by each contributing factor, we perform ablation studies.

Results

Model accuracy

Table 1 details the absolute errors in aortic area (in mm²) and AD (in mmHg⁻¹) as well as the Dice coefficient for both AAO and DAo. The proposed method attained lower absolute area and AD errors and higher Dice coefficient values in comparison to both the state-of-the-art (SOTA) and unpruned methods. We have observed every model reaching a plateau before 250 epochs. Figures 1, 2 and 3 display the (maximum and minimum) area and AD Bland-Altman analysis plots for the proposed, SOTA and unpruned methods, respectively, versus the ground truth. It can be seen from the the limits of agreement of the Bland-Altman analysis plots versus the ground truth that the maximum and minimum aorta area as well as the AD values predicted by the proposed method had, on average, ~ 3.9 times less fluctuation than the SOTA and unpruned methods. Figure 4 qualitatively compares the proposed and SOTA networks for three cases. Case 1 shows an instance where the SOTA method severely underestimated the number of pixels that had to be assigned to the aorta. Case 2 is a patient where the SOTA method over-segmented the AAO. Case 3 is a rare instance where the DAo is in an abnormal position relative to the AAO and for which the SOTA method was unable to identify the AAO. Our method accurately segmented both AAO and DAo for all three cases (Table 2).

Table 3 shows the results of the qualitative and quantitative comparisons of ascending and descending aorta time-area curves for one cardiac cycle of a representative case. It is apparent that the cardiac cycle phases of the proposed and the ground truth curves are in much closer agreement (for both AAO and DAo) than in the SOTA cases.

Model resource efficiency

Table 2 gives the predicted carbon dioxide equivalent (CO₂eq) emissions (in g), the consumed energy (in kWh) and the equivalent distance (in km) a car could travel during the final training (250 epochs) for the proposed, SOTA and unpruned models. Also listed in Table 2 are the training time (in h:min:s) and the average inference

Model	Absolute error in area (mm ²)		Absolute error in AD (10 ⁻³ × mmHg ⁻¹)		Dice coefficient	
	AAo	DAo	AAo	DAo	AAo	DAo
Proposed, mean (± SD)	7.346 (± 2.257)	4.749 (± 2.567)	0.394 (± 0.401)	0.544 (± 0.908)	0.989 (± 0.003)	0.991 (± 0.004)
SOTA ²⁴ , mean (± SD)	32.323 (± 25.420)	11.809 (± 10.204)	1.088 (± 1.395)	0.942 (± 2.013)	0.965 (± 0.0161)	0.978 (± 0.015)
Unpruned ³¹ , mean (± SD)	30.739 (± 21.110)	12.389 (± 7.209)	2.490 (± 2.218)	1.404 (± 1.759)	0.980 (± 0.009)	0.970 (± 0.012)
Proposed Vs SOTA ²⁴ <i>p</i> values	1.710 × 10 ⁻¹⁵	1.539 × 10 ⁻¹⁴	5.291 × 10 ⁻⁰⁸	9.022 × 10 ⁻⁰³	1.710 × 10 ⁻¹⁵	1.710 × 10 ⁻¹⁵
Proposed Vs Unpruned ³¹ <i>p</i> values	1.711 × 10 ⁻¹⁵	3.503 × 10 ⁻¹⁵	7.427 × 10 ⁻¹³	4.888 × 10 ⁻⁰⁷	1.651 × 10 ⁻¹⁴	1.779 × 10 ⁻¹⁵

Table 1. Quantitative end-to-end performance of the proposed method using the absolute errors in aortic area and aortic distensibility (AD) and the Dice coefficient. Also provided are the *p* values obtained from the Wilcoxon-signed rank test with Bonferroni correction ($\alpha = 0.05$). The mean ground truth area (averaged over time and patients) was 678.826 mm² (SD: 146.329 mm²) for the AAO and 370.610 mm² (SD: 85.109 mm²) for the DAo. Bold face indicates best performance.

Model	CO ₂ eq (g)	Energy (kWh)	Equivalent distance travelled by car (km)	Training time (h:min:s)	Average inference time (ms)
Proposed	2093.571	5.984	17.388	06:46:11	2.768
SOTA ²⁴	5785.498	23.184	48.052	35:09:20	7.544
Unpruned ³¹	11031.780	39.574	91.626	63:06:35	12.25

Table 2. Resource efficiency evaluation of the proposed method using the generated carbon emissions, the consumed energy and the equivalent distance a car could travel during the final training (250 epochs). Also shown are the training and average inference times. The experiments were conducted on a workstation with an NVIDIA RTX A6000 (48GB) GPU. Bold face indicates best performance.

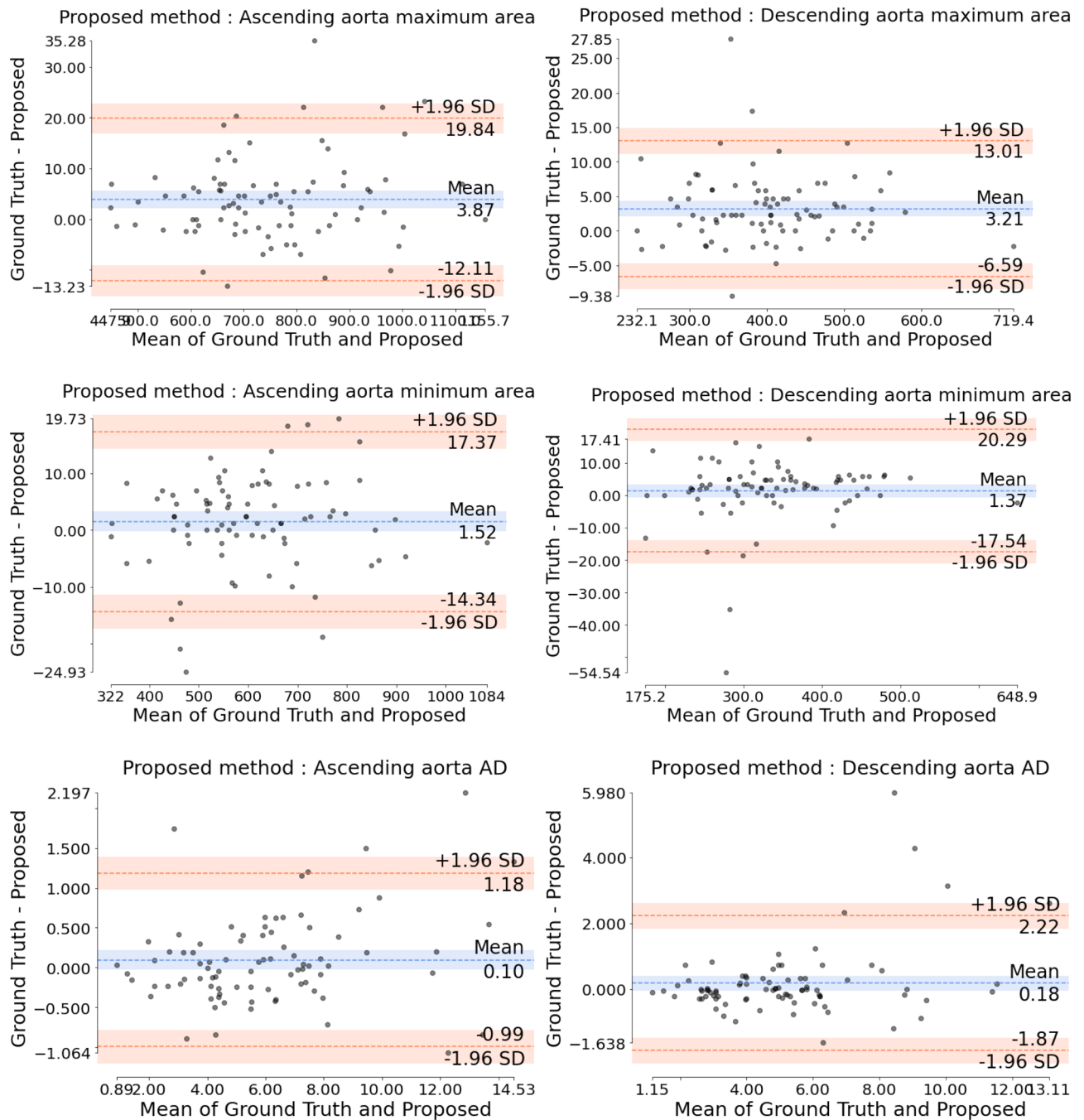


Figure 1. Bland-Altman analysis for graphically comparing the proposed method to the ground truth with respect to aorta maximum areas, aorta minimum areas and aortic distensibility (AD) values. Y-axis gives the difference between the two methods whereas X-axis represents their mean. Area is measured in mm^2 . AD is measured in $10^{-3} \times \text{mmHg}^{-1}$. SD is the standard deviation.

time (in ms). The proposed method was ~ 2.8 times more environmentally friendly and it consumed ~ 3.9 times less fuel when compared with SOTA. In addition, it required ~ 5.2 times less time to train, and its inference time was ~ 2.7 times faster. Our method was also ~ 5.2 times less polluting and it consumed ~ 6.6 times less fuel when compared with the unpruned method. Lastly, it required ~ 9.3 times less time to train, and its inference time was ~ 4.4 times faster than the unpruned version.

Ablation study

In this section, we carried out ablation experiments over a number of features of the proposed framework to better appreciate their proportionate significance. In particular:

No full labels: is trained using labels obtained by propagating ES and ED labels using non-rigid image registration.

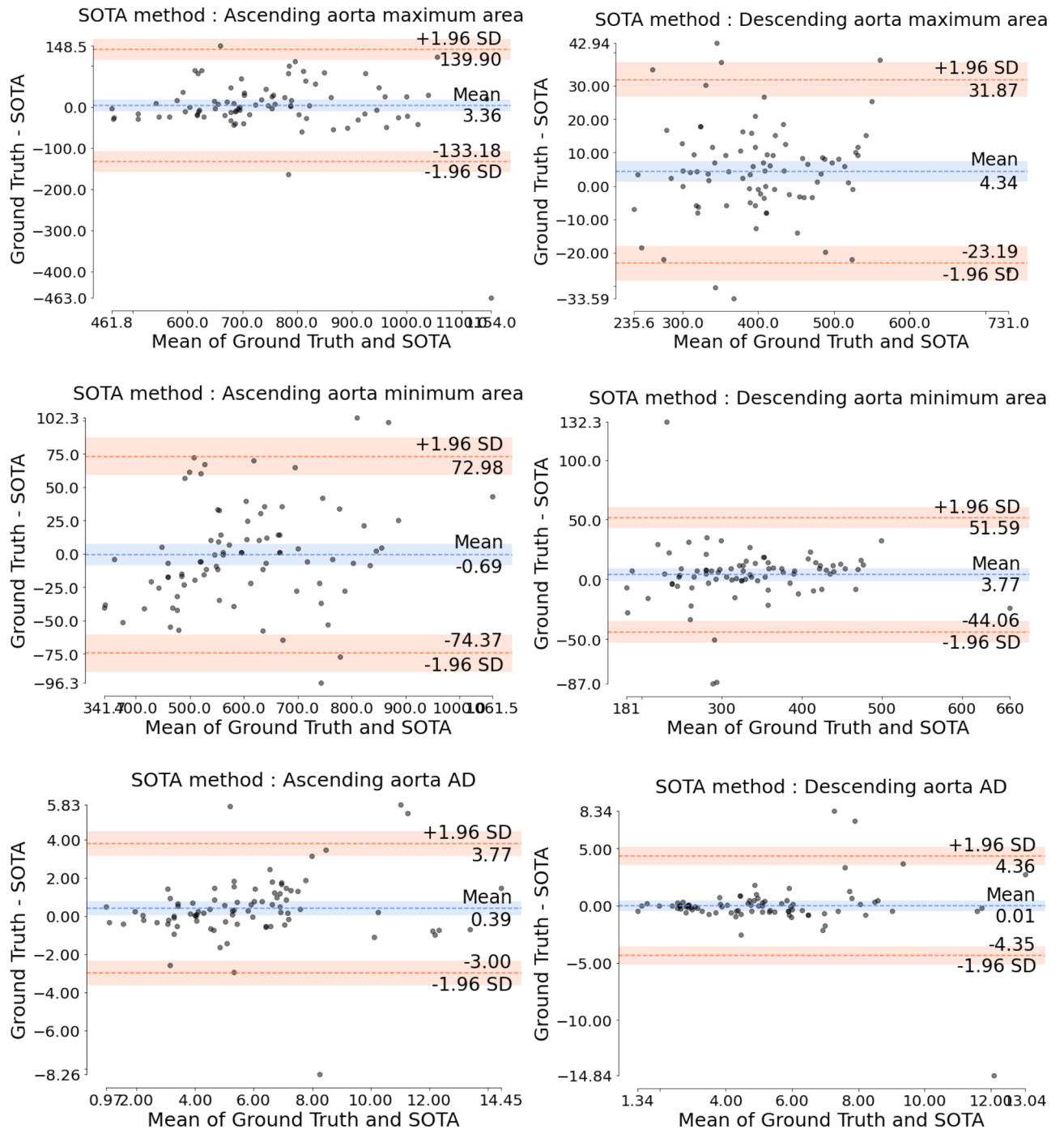


Figure 2. Bland-Altman analysis for graphically comparing the SOTA method²⁴ to the ground truth with respect to aorta maximum areas, aorta minimum areas and aortic distensibility (AD) values. Y-axis gives the difference between the two methods whereas X-axis represents their mean. Area is measured in mm². AD is measured in 10⁻³ × mmHg⁻¹. SD is the standard deviation.

- No focal Tversky:* is trained using Dice coefficient loss, ignoring the severe class imbalance in the data.
- No non-linearities:* is trained by merging the encoder and decoder feature maps through linear concatenation.
- No dense pruning:* is trained using three densely packed convolutional blocks in the final encoding step.
- No filter pruning:* is trained using four times more filters in the convolutional layers than the proposed model.
- No BConvLSTM pruning:* is trained using BConvLSTM unit in three steps instead of one.

To explore the effect of each factor on model accuracy, we employed the same evaluation metrics that were used in Table 1 and the same clinical images and hyperparameters that were used in the proposed framework. Table 4 gives the ablation results. It can be seen that not using non-linearities caused the largest drop in AD calculation performance. It is surprising that increasing the model size did not lead to accuracy improvement. This was possibly due to overfitting. In addition, not using focal Tversky loss and the full labels had more significant

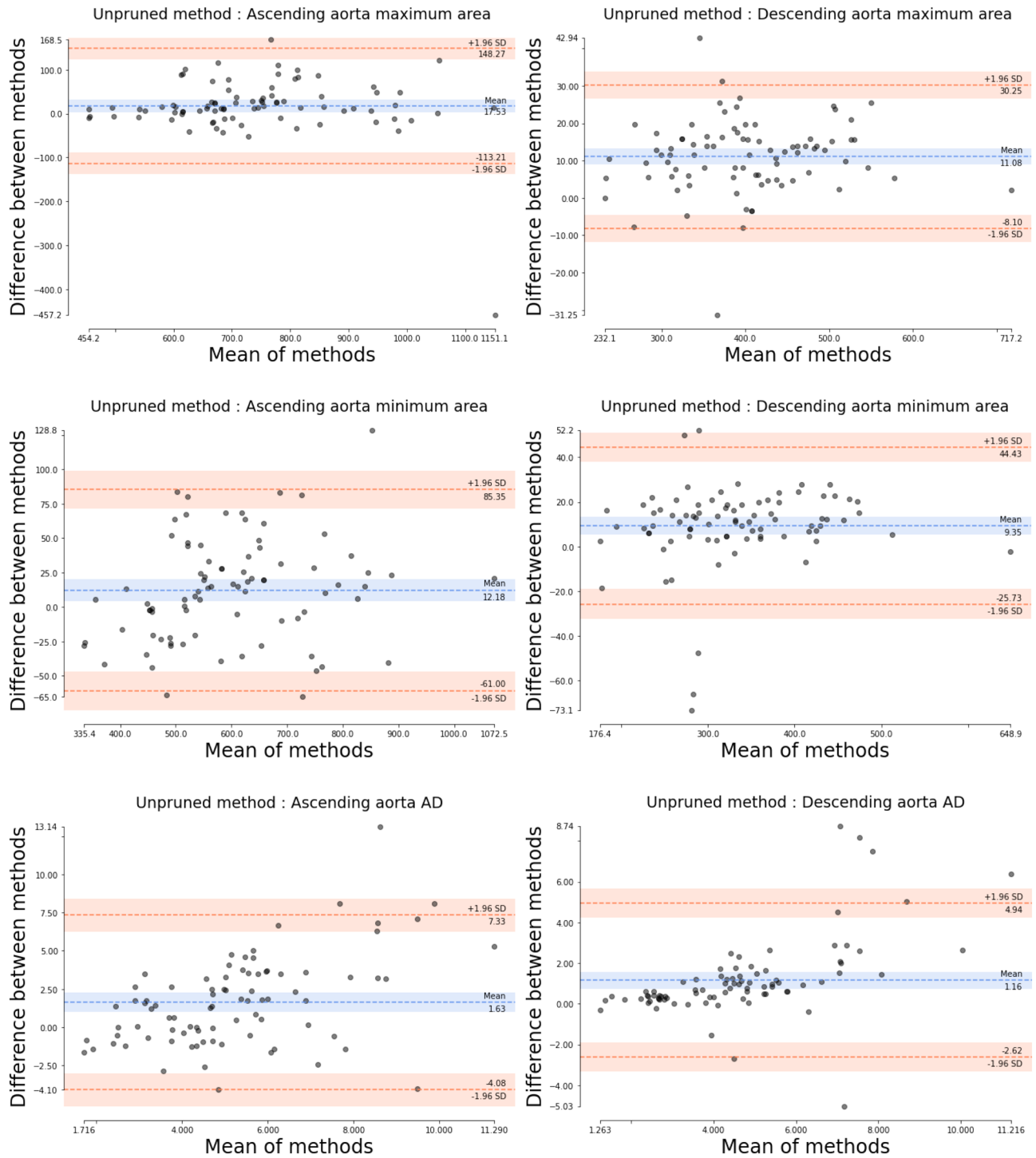


Figure 3. Bland-Altman analysis for graphically comparing the unpruned method³¹ to the ground truth with respect to aorta maximum areas, aorta minimum areas and aortic distensibility (AD) values. Y-axis gives the difference between the two methods whereas X-axis represents their mean. Area is measured in mm². AD is measured in 10⁻³ × mmHg⁻¹. SD is the standard deviation.

impact on the AAO AD calculation, most likely due to the increased number of false positives caused by surrounding structures. Lastly, it can be seen that the drop in performance for both the absolute error in area and Dice coefficient are large and uniform for every row in Table 4, whereas the absolute error in AD is a lot more variable. The reason for this is that the AD calculation is more susceptible to maximum and minimum area segmentation artefacts, while the segmentation accuracy metrics are averaged over the whole cardiac cycle, thus dampening the artefacts' impact.

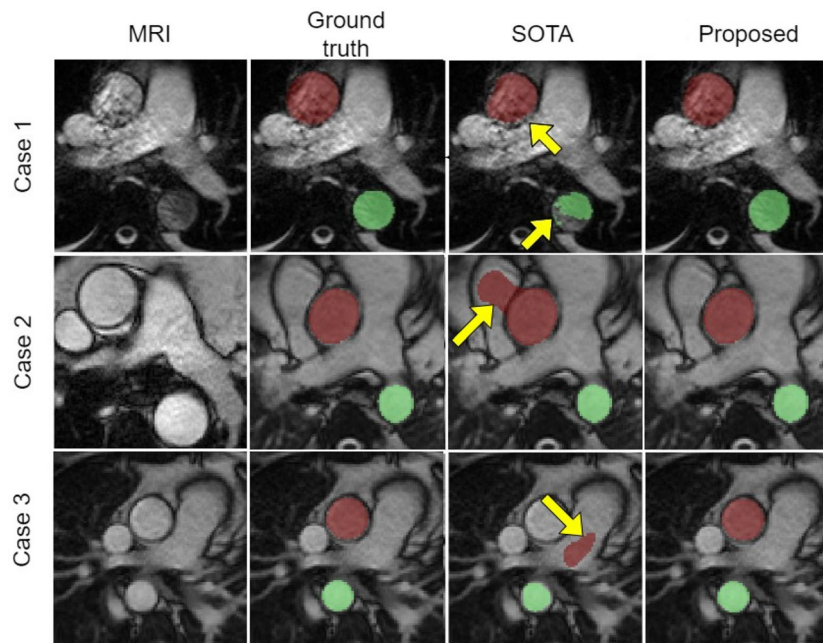


Figure 4. Qualitative comparison of the proposed method with the state-of-the-art (SOTA) method²⁴ in one time frame of the cardiac cycle for three representative cases. First column: MRI. Second column: Ground truth (semi-automated segmentation). Third column: Segmentation results of the SOTA method²⁴. Fourth column: Segmentation results of the proposed method. The yellow arrows indicate the errors in segmentation. The ascending aorta (AAo) is denoted by the red area and the descending aorta (DAo) is denoted by the green area.

AAo is the ascending aorta, DAo is the descending aorta, SD is the standard deviation and SOTA represents the state-of-the-art method²⁴.

SOTA denotes the state-of-the-art method²⁴.

Ethical approval and consent to participate

Each study was approved by the UK national research and ethics service and written informed consent was obtained from all subjects prior to participation.

Discussion

The elevated aortic stiffness has been linked to an increased risk of cardiovascular diseases. In this study, a novel resource-efficient DL model was proposed for the fast, robust, and fully automated segmentation of the AAo and the DAo from aortic cine CMR images towards streamlining quantification of AD.

Strengths of this study

This study has a number of strengths compared to previous work. We resolved to not apply the semantic segmentation algorithm to each time frame of the cardiac cycle as a separate entity. Instead, we incorporated information related to both space and time into our task by merging the encoder and decoder feature maps in the second up-sampling layer through a non-linear function, namely a BConvLSTM. The use of the hyperbolic tangent function for combining the output of the forward and backward paths helps the network learn complex data structures³¹. To address the issue that the foreground is significantly smaller than the background class, we employed the focal Tversky loss during training. Following the enhanced credibility of the ground truth targets in all cardiac cycle time frames, we performed end-to-end hierarchical learning and testing of the aortic lumen area from cine CMR images, rather than focusing on a very limited range of time frames in the cardiac cycle. Another strength of this study is that we used multi-centre multi-vendor data from a highly diverse patient cohort which significantly adds to the ability of the proposed aortic lumen segmentation algorithm to generalise. The site, vendor, and patient heterogeneity when testing a model are crucial to the effective clinical implementation and to get approval by accreditation agencies. Unlike the model that motivated this paper³¹, our algorithm uses a building block that returns the sequence of feature maps over all time steps since we are dealing with video inputs. In addition, and in pursuit of a resource-efficient architecture, our algorithm: (i) has four times fewer filters in the convolutional layers, (ii) involves BConvLSTM layers in three times fewer steps, and (iii) contains only one third of densely packed convolutional blocks in the final encoding step.

Main findings

The AAo and DAo segmentation masks predicted by our model were in close agreement with the semi-automated annotations by the CMR experts, as evidenced by the low area errors and the high Dice coefficient values. This

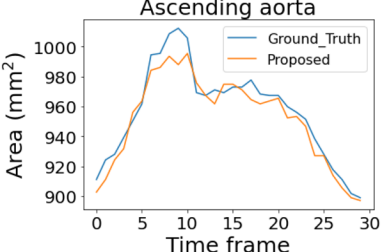
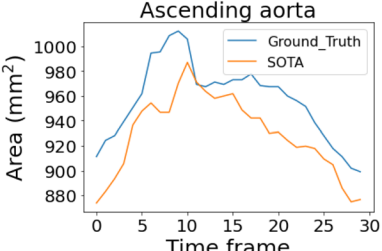
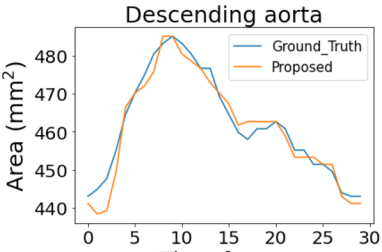
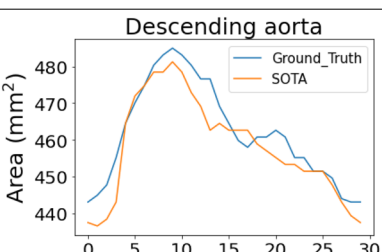
Temporal curve	Fréchet distance	Hausdorff distance	Dynamic time warping (DTW) distance
<p>Ascending aorta</p> 	16.822	16.822	165.424
<p>Ascending aorta</p> 	37.384	37.384	337.390
<p>Descending aorta</p> 	5.607	4.673	63.552
<p>Descending aorta</p> 	9.346	6.542	85.048

Table 3. Qualitative and quantitative comparisons of ascending and descending aorta time-area curves for one cardiac cycle of a representative case. SOTA denotes the state-of-the-art method²⁴. Bold face indicates best performance.

finding corroborates that only one densely packed convolution block in the final layer of the contracting path sufficed to learn diverse features³⁵. The proposed method was compared to other DL models that represent the SOTA and unpruned methods^{24,31}. The quantitative analysis showed that our model outperformed both SOTA and unpruned methods in terms of segmentation accuracy, attaining lower aortic area and AD absolute errors and higher Dice coefficient values for both AAO and DAO. Furthermore, the limits of agreement of the Bland-Altman analysis plots versus the ground truth revealed that the maximum and minimum aorta area as well as the AD values predicted by the proposed method had, on average, ~ 3.9 times less fluctuation than the SOTA and unpruned methods. These findings are in keeping with recent studies³⁶, which reported that more powerful class-discriminative features can be captured in an efficient manner by tying up the temporal with the spatial processing in video inputs. Importantly and as an indication of an enhanced ability to generalise, our method, unlike SOTA, was found to accurately segment the aortic areas for a rare (and unseen by the algorithm) case for which the DAO was located at an abnormal position with respect to the AAO. Such an infrequent condition can be observed in individuals with scoliosis or patients with aortic arch branching and orientation abnormalities. Unlike previous studies^{24,25}, the performance of the evaluated models on the AAO was slightly inferior to that on the DAO. A reason for this observation might be that the structures surrounding the AAO seem to be more challenging to differentiate between them in our diverse dataset. The consistency of the proposed segmentation method across the cardiac cycle was double-checked by demonstrating AAO and DAO time vs. area curves and estimating curve distances from the ground truth for a representative case. Finally, the proposed method was compared to SOTA and unpruned methods also in terms of resource-efficiency and carbon footprint using the

Model	Absolute error in area (mm ²)		Absolute error in AD (10 ⁻³ × mmHg ⁻¹)		Dice coefficient	
	AAo	DAo	AAo	DAo	AAo	DAo
Proposed, mean (± SD)	7.346 (± 2.257)	4.749 (± 2.567)	0.394 (± 0.401)	0.544 (± 0.908)	0.989 (± 0.003)	0.991 (± 0.004)
No full labels, mean (± SD)	29.964 (± 25.754)	8.649 (± 7.776)	4.017 (± 3.230)	3.038 (± 2.581)	0.983 (± 0.011)	0.970 (± 0.016)
No focal Tversky, mean (± SD)	30.034 (± 28.886)	10.399 (± 6.815)	25.499 (± 148.856)	2.161 (± 1.807)	0.981 (± 0.009)	0.969 (± 0.012)
No non-linearities, mean (± SD)	30.589 (± 20.173)	10.367 (± 7.943)	30.761 (± 210.163)	49.617 (± 437.580)	0.980 (± 0.016)	0.967 (± 0.0154)
No dense pruning, mean (± SD)	30.034 (± 28.886)	10.399 (± 6.815)	25.483 (± 148.859)	2.134 (± 1.86)	0.981 (± 0.008)	0.969 (± 0.012)
No filter pruning, mean (± SD)	28.966 (± 28.717)	9.322 (± 6.152)	10.901 (± 58.238)	1.991 (± 1.701)	0.982 (± 0.009)	0.971 (± 0.012)
No BConvLSTM pruning, mean (± SD)	46.811 (± 64.332)	17.288 (± 8.793)	46.502 (± 274.518)	2.116 (± 2.227)	0.971 (± 0.011)	0.962 (± 0.014)

Table 4. Ablation over number of features using the proposed framework. “No full labels” is trained using labels obtained by propagating ES and ED labels using non-rigid image registration. “No focal Tversky” is trained using Dice coefficient loss, ignoring the severe class imbalance in the data. “No non-linearities” is trained by merging the encoder and decoder feature maps through linear concatenation. “No dense pruning” is trained using three densely packed convolutional blocks in the final encoding step. “No filter pruning” is trained using four times more filters in the convolutional layers than the proposed model. “No BConvLSTM pruning” is trained using BConvLSTM unit in three steps instead of one. Bold face indicates best performance.

Carbontracker tool³⁷. Our method was found to consume ~ 3.9 times less fuel and it was ~ 2.8 less polluting during training than SOTA. In addition, it required ~ 5.2 times less time to train, and its inference time was ~ 2.7 times shorter than SOTA. As expected, our method was also ~ 5.2 times less polluting and it consumed ~ 6.6 times less fuel when compared with the unpruned method. Lastly, it required ~ 9.3 times less time to train, and its inference time was ~ 4.4 times faster than the unpruned version. Ablation studies showed that using BConvLSTM non-linearities gave the largest boost in AD calculation accuracy. Increasing the number of model parameters led to overfitting. Lastly, not using focal Tversky loss and the full labels had more significant impact on the AAo AD calculation, most likely due to the increased number of false positives caused by surrounding structures.

Clinical implications

Previous experiments conducted in our lab have shown that the AD aortic stiffness measure has better reproducibility than the pulse wave velocity³⁸. The excellent results presented in this paper allow to significantly reduce the time spent on extracting aortic structural and functional phenotypes from CMR data, while also improving the reliability of the results. Therefore, our pipeline could provide a valuable resource for exploring genome-wide associations of the AD and aortic areas with the cognitive performance for very large-scale biomedical databases (such as the UK Biobank)^{39,40}, which better represent the wider population. Obtaining quantitative CMR phenotypes at such a scale remains a challenge nowadays. This type of analysis would enable us to investigate possible causal relationships of the aortic measures with aortic aneurysms and brain small vessel disease, as well the bidirectional relationship with blood pressure indices. The definition of the responsible mechanisms would eventually lead to (i) improved understanding of the factors that contribute to cognitive decline and dementia, and (ii) identification of new therapeutic targets. Moreover, the model developed for the AD quantification task could be reused as the starting point for various clinically relevant tasks (through transfer learning) to allow rapid progress and enhanced performance. The proposed framework could easily be integrated within a CMR analysis software. We have shared the code of our image analysis pipeline online (<https://github.com/tuanaqeelb/ohoran/Aortic-Distensibility.git>).

For our method to substantially impact routine clinical care, appropriate infrastructure for deployment and evaluation within the healthcare centre is needed. This ideally should include ML-ops platforms, pipelines for ensuring safety standards through identifying potential (such as methods for recognizing when the algorithm is operating in a more complex landscape than the one used for training or model explainability methods that highlight decision-relevant parts of feature representations).

Resource efficiency considerations

It was found that the computations required for DL research have grown 300,000-fold from 2012 to 2018⁴¹, which is much faster than the rate at which compute demand has historically increased. These computations require staggering amounts of energy for fuelling them³². Given that electricity usage is correlated with greenhouse gas emissions, the associated carbon footprint is also outsized which appears to accelerate global climate change³². Even though the biggest part of this problem is caused by the very large-scale models used in natural language processing applications, the energy consumed by typical medical image analysis DL models is far from negligible. We estimated that the energy required to train our model was 5.984 KWh, as opposed to 23.183 KWh for the SOTA model. The respective generated carbon emissions were equivalent to those produced when a car travels a distance of 17.388 km for our model versus 48.052 km for SOTA. However, it is important to understand that those figures refer to the final training only. Before the final training run, resolving the optimal model in a

standard development process necessitates multiple training runs due to the need for hyperparameter tuning and experimenting with various model architectures. To get a better handle on what the full DL model development pipeline might look like in terms of energy usage and carbon footprint, a recent study found that the process of building and testing a final paper-worthy DL model required training 4789 models over a 6-month period⁴². The significance of the dire numbers mentioned above is colossal, especially after taking the worldwide adoption of healthcare DL applications into consideration. Despite the fact that few concerns involving the energy usage and carbon footprint of DL research started to emerge^{32,37,42}, the overwhelming majority of DL research in computer vision for healthcare are only concerned with enhancing accuracy while ignoring resource efficiency. In this paper, we made energy usage and CO₂e emissions explicit. We endorse prior calls for making those key metrics evaluation criteria alongside accuracy-related measures in DL research, as a means to accelerate innovations in DL algorithmic efficiency^{32,42–45}. The value of DL models should be judged by the amount of intelligence they provide per joule. Such an initiative would ultimately help: (i) reduce the adverse environmental impact of DL research during training and development, and (ii) make DL research more inclusive by letting more people participate⁴⁵. With the spectre of an energy-hungry future looming alongside the escalating rates of natural disaster, it is imperative to explore ways to keep DLs energy usage and carbon cost in check³². Finally, it is worth noting that the resource efficiency element is appealing also because it might increase the portability and practicality of our approach by promoting its wider distribution to devices with lower computational power and memory, which in turn might open up the more universal and systematic DL-powered evaluation of the CMR-derived aortic stiffness.

Study limitations

This study has several limitations. Despite the large and diverse training dataset, our network is likely to produce less-accurate results if it is presented with pathologies, ages, ethnic backgrounds, and scanners that have not been included in the training set. In fact, this is the main limitation hampering the deployment of any DL model in the real world. To deal with this issue to a certain extent, we employed data augmentation techniques that simulate various possible data distributions. Altogether, it was illustrated that our pipeline exhibited improved ability to generalise to an unseen non-representative case compared to SOTA. Another main limitation of our model is that it lacks interpretability due to the intrinsic “black-box” nature of the DL algorithms. Explainable tools are a prerequisite for building trust in DL models, and the development of such a module will be a topic of future research⁴⁶. Furthermore, our model has not been tested against adversarial attacks⁴⁷. Even though Carbontracker supports a variety of different environments and platforms, there might be small deviations of the reported energy consumption and carbon emission values from the true ones. Such deviations might be caused by the quality of the estimated carbon intensity of electricity production which varies by geographic location (depending on the energy sources that power the local electrical grid) and time throughout the day (as energy demand and capacity change). However, all DL experiments in this study were conducted at the same workstation and time period during the day. In addition, Carbontracker uses “real-time” carbon intensity values which are fetched every 15 min during training using the application programming interfaces supported by this tool. In this study, the predicted maximum and minimum aortic areas, that were also used for the quantification of AD, did not always relate with the diastolic and systolic cardiac cycle phases, respectively, of the handcrafted analysis. However, this did not significantly affect the AD measurements. Our study required full annotation of all temporal frames in the cine dataset, as opposed to SOTA that required only sparse annotation. However, the SOTA method also needs extra resources (people, time) to identify ED and ES phases. It also needs significant resources (hardware, time) to perform the very computationally-intensive non-rigid registration. Lastly, the annotation in our study was not so time-consuming because it was largely supported by the JIM software that performed automated label propagation, which was then verified by the clinicians. Therefore, the consensus on the standards is that our study is concerned only with training resources. Finally, other recent fully-automated segmentation approaches do exist^{48,49}, that could potentially perform better than ours in the aortic lumen delineation task from cine CMR images. However, the goal of this work was to propose a DL-based framework that surpasses the SOTA methods for this particular task while being more resource-efficient, rather than to carry out an exhaustive survey of semantic segmentation, the literature of which is huge.

Conclusion

This paper proposed a novel resource-efficient DL model for the fast, robust, and fully-automated segmentation of the AAo and DAo from aortic cine CMR images towards streamlining quantification of AD. When evaluated on a large multi-centre multi-vendor dataset from a highly heterogeneous patient cohort, the proposed method outperformed the SOTA method in terms of accuracy and at the same time it consumed ~ 3.9 times less fuel and generated ~ 2.8 less carbon emissions. Notably, the proposed method was even more accurate than the unpruned method. Our model could provide a valuable tool for exploring genome-wide associations of the AD and aortic areas with the cognitive performance in very large-scale biomedical databases. By making energy usage and greenhouse gas emissions explicit, the presented work aligns with efforts to keep DLs energy requirements and carbon cost in check. The improved resource efficiency element of our pipeline might open up the more universal and systematic DL-powered evaluation of the CMR-derived aortic stiffness.

Methods

Study population and image dataset

The study population comprises participants from four clinical studies analysed at the University Hospitals of Leicester NHS Trust MRI core lab which included AD assessment. These included participants with spontaneous coronary artery dissection⁵⁰, asymptomatic type 2 diabetes (from Lydia⁵¹ and DIASTOLIC⁵² trials), hypertension

(the Pathway 2 study) and healthy volunteers (recruited in above studies^{50,51}). In total, we analysed 424 aortic MRI datasets taken from 376 patients. The demographic, anthropometric and clinical characteristics of the participants are presented in Table 5. Each study was approved by the UK national research and ethics service and written informed consent was obtained from all subjects prior to participation. All methods were performed in accordance with the relevant guidelines and regulations.

MRI acquisition

Acquisitions were undertaken on either 1.5T or 3T scanners from three centres using four different MRI scanners: Leicester (Siemens Aera, 1.5T and Skyra, 3T), Cambridge (GE Signa, 1.5T) and Dundee (Siemens TrioTim, 3T). SSFP cine images of the AAO and DAO in a plane perpendicular to the thoracic aorta at the level of the pulmonary artery bifurcation were reconstructed to 30–40 phases as previously described^{20,38}. The typical image matrix was 256 × 186 to 256 pixels. The in-plane pixel height and width varied between 1.093mm and 1.914mm. The slice thickness for cine imaging was 8mm. Brachial blood pressure was measured simultaneously to determine pulse pressure.

Data pre-processing and annotation

The end-to-end data annotation was carried out semi-automatically by experts from the Glenfield hospital in Leicester using the Java Image Manipulation Software Version 6 (Xinapse Systems Ltd, Essex, UK)⁵³ as previously described^{20,38}. The ascending and descending aorta was analysed by manually contouring the first and last phase and every 6th phase in between. Then images were propagated to fill in the other phases of the sequence. All phases were then manually checked and adjusted if required. The experts were blinded to the patient details. All MRI images and masked images (annotations) were zero-padded to 256 × 256 pixels so that their dimensions match.

Neural network architecture

The proposed architecture (Fig. 5) was inspired by the BConvLSTM U-Net with densely connected convolutions³¹. It consists of: (i) a contracting path (a.k.a. the encoder) for capturing the context in the image by transforming it into a high-level feature representation, and (ii) a symmetric expanding path (a.k.a. the decoder) for interpreting the feature maps, enabling precise localisation (where in the image) and producing a full resolution segmentation map. There are four down-sampling layers in the encoder and four up-sampling layers in the decoder.

Each encoding step consists of a sequence of two convolutional layers (3 × 3 filters and a ReLU non-linear activation⁵⁴) followed by a Batch Normalisation (BN) layer⁵⁵ and a 2 × 2 max-pooling. BN was employed so that the optimiser converges faster when training the network, whereas the max-pooling operation was used to aggressively down-sample the feature maps. To reduce over-fitting, dropout regularisation⁵⁶ was employed in the last two steps of the contracting path before the BN layer. The number of filters that each encoding layer computes over the input doubles at each step ([16, 32, 64, 128, 256]).

Each decoding layer consists of a sequence of two convolutional layers (3 × 3 filters and a ReLU non-linear activation) apart from the final decoding step that has five convolutional layers. In each step of the expanding path, the output of the previous layer is passed onto an up-conv layer (i.e. up-sampling function followed by a 2 × 2 convolution; this process doubles the size of the feature map) and then combined with the corresponding (same-resolution-level) representation in the contracting path using skip connections. The combination of these two types of feature maps is a channel-wise concatenation in all steps except for the second up-sampling layer, where we propose to merge them in a more complex way using (apart from concatenation also) a BConvLSTM³³ building block that outputs information about all temporal hidden states. This is to account for the spatiotemporal composition of the input. The BConvLSTM building block is made of two ConvLSTMs. ConvLSTM is a variant of the traditional LSTM neural network that is specifically designed to process spatial data. By incorporating convolutional structures within the LSTM gates, ConvLSTM is capable of capturing spatial dependencies in the data. Figure 5 provides a visual representation of the ConvLSTM block operation. The ConvLSTM block's

	(n = 376)
Age, mean (± SD), y	48 (± 8)
Male, No. (%)	150 (40)
Female, No. (%)	226 (60)
BMI, mean (± SD), kg/m ²	31.74 (± 7.41)
SBP, mean (± SD), mmHg	127.33 (± 17.44)
DBP, mean (± SD), mmHg	79 (± 12.59)
Hypertension, No. (%)	197 (52)
Smoking, No. (%)	192 (51)
History of CVA, No. (%)	2 (1)
History of DM, No. (%)	208 (55)
Renal impairment, No. (%)	78 (21)

Table 5. Patient characteristics. Abbreviations: BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; CVA, cerebrovascular accident; DM, type 2 diabetes mellitus.

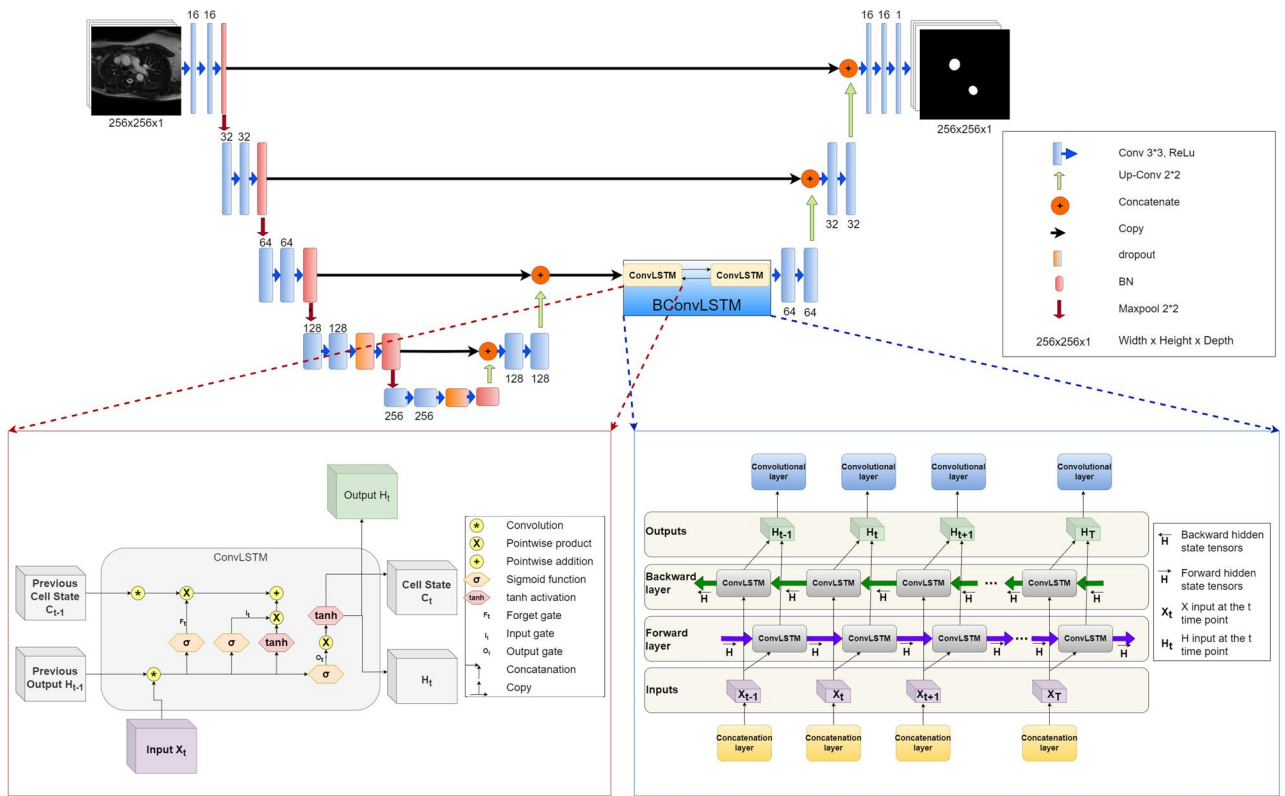


Figure 5. The proposed deep learning model, including two diagrams that illustrate how the temporal dimension is handled.

distinctiveness lies in its ability to apply convolutional operations within the gates, enabling the capture of spatial dependencies in the data. Then, the BCConvLSTM output Y_j at the time step j is calculated as

$$Y_j = \tanh(W_y^{\vec{H}} * \vec{H}_j + W_y^{\overleftarrow{H}} \overleftarrow{H}_j) + b, \tag{2}$$

where \vec{H} and \overleftarrow{H} denote forward and backward hidden state tensors³¹, respectively, $W_y^{\vec{H}}$ and $W_y^{\overleftarrow{H}}$ denote the forward and backward convolution kernels corresponding to the hidden states, b represents the bias term, and the hyperbolic tangent was employed to combine the outputs of the forward and backward paths in a non-linear way. The mathematical equations for obtaining each of the two hidden state tensors are as follows:

$$I_t = \sigma(W_{XI} * X_t + W_{HI} * H_{t-1} + b_I) \tag{3}$$

$$F_t = \sigma(W_{XF} * X_t + W_{HF} * H_{t-1} + b_F) \tag{4}$$

$$C_t = F_t \circ C_{t-1} + I_t \circ \tanh(W_{XC} * X_t + W_{HC} * H_{t-1} + b_C) \tag{5}$$

$$O_t = \sigma(W_{XO} * X_t + W_{HO} * H_{t-1} + b_O) \tag{6}$$

$$H_t = O_t \circ \tanh(C_t) \tag{7}$$

Here, I_t , F_t , and O_t denote the input, forget, and output gates, respectively, at time t , while C_t and σ represent the cell state and the sigmoid function, respectively, and x_t refers to the input at time t (Fig. 5). The input gate governs the information that is retained in the cell state. The forget gate regulates the information that is discarded from the cell state. The output gate controls the information utilized to compute the output of the LSTM. The cell state maintains the internal state of the ConvLSTM. The $*$ operator symbolises the convolution operation, and \circ represents the Hadamard product (element-wise multiplication).

The number of channels reduces in every step of the expanding path ([256, 128, 64, 32, 16, 1]), whereas the size of the feature maps progressively increases to reach the input size after the final layer.

Unlike the Azad et al.³¹ that inspired this work, our method uses a BCConvLSTM building block that returns the sequence of feature maps over all time steps since the dataset was time distributed. In addition, and in pursuit of a more efficient architecture, our method: (i) uses four times less filters in the convolutional layers compared to Azad et al.³¹, (ii) involves BCConvLSTM in only one step, (iii) contains only one densely packed convolutional block in the final encoding step.

Implementation and training

The dataset was randomly split into a training set (272 datasets), validation set (68 datasets) and a testing set (84 datasets). For training, we used the Adam optimiser⁵⁷ for 250 epochs with a constant learning rate of 0.001 and a batch size of 120 (approximately 4 patients). The dropout value that we used was 0.5. The initialiser of the network was He Normal⁵⁸. In order to improve the proposed model's ability to generalise, online data augmentation techniques were employed. The augmented data were obtained from the original images by applying random rotations (by a degree between -30° and $+30^\circ$) and random translations along the x- and/or y-axis in either direction (by up to 20 pixels). All hyperparameters were tuned using grid search based on the validation accuracy. To address the severe class imbalance between pixel values 0 and 1 in each frame, we utilised the Focal Tversky loss function defined as

$$\text{Focal Tversky Loss} = (1 - \text{Tversky Loss})^\gamma \quad (8)$$

where γ is the adjustable focussing parameter and Tversky Loss is the Tversky index given by

$$\text{Tversky Loss} = \frac{\sum_i^N p_{0i}g_{0i}}{\sum_i^N p_{0i}g_{0i} + \alpha \sum_i^N p_{0i}g_{1i} + \beta \sum_i^N p_{1i}g_{0i}} \quad (9)$$

where p_{0i} is the probability that pixel i belongs to the aorta and p_{1i} is the probability of pixel i being in the background class^{59,60}. In addition, g_{0i} is 1 for the aortic vessel segmentation area and 0 for the background, and the opposite is true for g_{1i} . Finally, α and β are variables in Tversky Loss, which control the magnitude of the penalties for false positives and false negatives, respectively. To improve model convergence and the recall rate^{59,60}, we trained our model with $\alpha = 0.8$, $\beta = 0.8$ and $\gamma = 1$. All the methods were trained using the TensorFlow environment.

Model evaluation and statistical analysis

For evaluating the automated segmentation masks produced by our method relatively to the ground truth, we employed the Dice coefficient as well as the absolute area error (in mm^2) and absolute AD error (in mmHg^{-1}). In addition, we used Bland-Altman analysis for assessing the agreement between (maximum and minimum) aorta areas and AD values⁶¹. The temporal fidelity of the segmentation performance across a cardiac cycle was assessed both qualitatively and quantitatively using the Fréchet, Hausdorff and dynamic time warping (DTW) distances for a representative case. To examine the impact brought by each contributing factor, we performed ablation studies. All analyses described above was performed for both the AAO and DAo. The reported results refer to the test set.

For evaluating the resource efficiency of our method, we calculated the CO_2 eq emissions (in g) and energy spent (in kWh) during training using the Carbontracker method³⁷. Carbontracker is an open-source tool written in Python for tracking and predicting the energy consumption and carbon emissions of training DL models for a given GPU. To put the carbon footprint produced during model training in context, we also reported the equivalent distance travelled by car⁶² that would generate the same emission volume. The training and inference times were also computed. For comparison purposes, all evaluation metrics of the proposed method were juxtaposed with those produced by the SOTA²⁴ and unpruned³¹ methods, also trained on the same multi-centre, multi-vendor multi-disease CMR dataset following similar training and hyperparameter tuning procedures as those described above. To determine if there is a statistically significant difference (at level 0.05) between the performances of the proposed and SOTA methods, we used the Wilcoxon signed-rank test with Bonferroni correction.

Data availability

The datasets generated and/or analysed during the current study are not publicly available due to privacy/ethical restrictions. If someone wants to request the data from this study, they should contact Prof. Gerry P. McCann at gpm12@leicester.ac.uk.

Code availability

The link to the GitHub repository containing the code of the image analysis pipeline is: <https://github.com/tuanaqelbohoran/Aortic-Distensibility.git>.

Received: 24 January 2023; Accepted: 2 December 2023

Published online: 08 December 2023

References

- London, G. M. & Guerin, A. P. Influence of arterial pulse and reflected waves on blood pressure and cardiac function. *Am. Heart J.* **138**, 220–224 (1999).
- Vaitkevicius, P. V. *et al.* Effects of age and aerobic capacity on arterial stiffness in healthy adults. *Circulation* **88**, 1456–1462 (1993).
- Cavalcante, J. L., Lima, J. A. C., Redheuil, A. & Al-Mallah, M. H. Aortic stiffness: Current understanding and future directions. *J. Am. Coll. Cardiol.* **57**, 1511–1522 (2011).
- Oxlund, H., Rasmussen, L. M., Andreassen, T. T. & Heickendorff, L. Increased aortic stiffness in patients with type 1 (insulin-dependent) diabetes mellitus. *Diabetologia* **32**, 748–752 (1989).
- Westenberg, J. J. M. *et al.* Age-related and regional changes of aortic stiffness in the Marfan syndrome: Assessment with velocity-encoded MRI. *J. Magn. Reson. Imaging* **34**, 526–531 (2011).
- Yasmin, *et al.* Variation in the human matrix metalloproteinase-9 gene is associated with arterial stiffness in healthy individuals. *Arterioscler. Thromb. Vasc. Biol.* **26**, 1799–1805 (2006).

7. Nistri, S. *et al.* Aortic elasticity and size in bicuspid aortic valve syndrome. *Eur. Heart J.* **29**, 472–479 (2007).
8. Chong, W.-Y., Wong, W. H. S., Chiu, C. S. W. & Cheung, Y.-F. Aortic root dilation and aortic elastic properties in children after repair of tetralogy of Fallot. *Am. J. Cardiol.* **97**, 905–909 (2006).
9. Redheuil, A. *et al.* Reduced ascending aortic strain and distensibility: Earliest manifestations of vascular aging in humans. *Hypertension* **55**, 319–326 (2010).
10. Cruickshank, K. *et al.* Aortic pulse-wave velocity and its relationship to mortality in diabetes and glucose intolerance: An integrated index of vascular function?. *Circulation* **106**, 2085–2090 (2002).
11. Willum-Hansen, T. *et al.* Prognostic value of aortic pulse wave velocity as index of arterial stiffness in the general population. *Circulation* **113**, 664–670 (2006).
12. Mattace-Raso, F. U. S. *et al.* Arterial stiffness and risk of coronary heart disease and stroke: The Rotterdam study. *Circulation* **113**, 657–663 (2006).
13. Redheuil, A. *et al.* Proximal aortic distensibility is an independent predictor of all-cause mortality and incident CV events: The MESA study. *J. Am. Coll. Cardiol.* **64**, 2619–2629 (2014).
14. O'Rourke, M. Mechanical principles in arterial disease. *Hypertension* **26**, 2–9 (1995).
15. Voges, I. *et al.* Normal values of aortic dimensions, distensibility, and pulse wave velocity in children and young adults: A cross-sectional study. *J. Cardiovasc. Magn. Reson.* **14**, 77 (2012).
16. Stefanadis, C., Stratos, C., Boudoulas, H., Kourouklis, C. & Toutouzas, P. Distensibility of the ascending aorta: Comparison of invasive and non-invasive techniques in healthy men and in men with coronary artery disease. *Eur. Heart J.* **11**, 990–996 (1990).
17. Dart, A. M. *et al.* Aortic distensibility in patients with isolated hypercholesterolaemia, coronary artery disease, or cardiac transplant. *Lancet* **338**, 270–273 (1991).
18. Resnick, L. M. *et al.* Direct magnetic resonance determination of aortic distensibility in essential hypertension: Relation to age, abdominal visceral fat, and in situ intracellular free magnesium. *Hypertension* **30**, 654–659 (1997).
19. Grotenhuis, H. B. *et al.* Validation and reproducibility of aortic pulse wave velocity as assessed with velocity-encoded MRI. *J. Magn. Reson. Imaging* **30**, 521–526 (2009).
20. Gulsin, G. S. *et al.* Relation of aortic stiffness to left ventricular remodeling in younger adults with type 2 diabetes. *Diabetes* **67**, 1395–1400 (2018).
21. Singh, A. *et al.* Aortic stiffness in aortic stenosis assessed by cardiovascular MRI: A comparison between bicuspid and tricuspid valves. *Eur. Radiol.* **29**, 2340–2349 (2018).
22. Herment, A. *et al.* Automated segmentation of the aorta from phase contrast MR images: Validation against expert tracing in healthy volunteers and in patients with a dilated aorta. *J. Magn. Reson. Imaging* **31**, 881–888 (2010).
23. van der Geest, R. J., Niezen, R. A., van der Wall, E. E., de Roos, A. & Reiber, J. H. Automated measurement of volume flow in the ascending aorta using MR velocity maps: Evaluation of inter- and intraobserver variability in healthy volunteers. *J. Comput. Assist. Tomogr.* **22**, 904–911 (1998).
24. Bai, W. *et al.* Recurrent neural networks for aortic image sequence segmentation with sparse annotations. In *Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, 2018, Proceedings. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (eds Frangi, A., Fichtinger, G., Schnabel, J., Alberola-López, C. & Davatzikos, C.) 586–594 (Springer-Verlag, Cham, 2018).
25. Hann, E. *et al.* Quality control-driven image segmentation towards reliable automatic image analysis in large-scale cardiovascular magnetic resonance aortic cine imaging. In *Medical Image Computing and Computer Assisted Intervention - MICCAI 2019* (eds Shen, D. *et al.*) 750–758 (Springer International Publishing, Cham, 2019).
26. Goodfellow, I. J., Bengio, Y. & Courville, A. *Deep Learning* (MIT Press, Cambridge, 2016).
27. Chollet, F. *Deep Learning with Python* 1st edn. (Manning Publications Co., New York, 2017).
28. Chen, C. *et al.* Deep learning for cardiac image segmentation: A review. *Front. Cardiovasc. Med.* **7**, 25 (2020).
29. Giannakidis, A. *et al.* Fast fully automatic segmentation of the severely abnormal human right ventricle from cardiovascular magnetic resonance images using a multi-scale 3d convolutional neural network. In *2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)* 42–46. <https://doi.org/10.1109/SITIS.2016.16> (2016).
30. Ras, G., Ambrogioni, L., Güçlü, U. & van Gerven, M. A. Temporal factorization of 3d convolutional kernels. arXiv preprint [arXiv:1912.04075](https://arxiv.org/abs/1912.04075) (2019).
31. Azad, R., Asadi-Aghbolaghi, M., Fathy, M. & Escalera, S. Bi-directional ConvLSTM U-Net with densely connected convolutions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops* (2019).
32. Schwartz, R., Dodge, J., Smith, N. A. & Etzioni, O. Green AI. *Commun. ACM* **63**, 54–63. <https://doi.org/10.1145/3381831> (2020).
33. Song, H., Wang, W., Zhao, S., Shen, J. & Lam, K. M. Pyramid dilated deeper convlstm for video salient object detection. In *Proceedings of the European Conference on Computer Vision (ECCV)* (2018).
34. Abraham, N. & Khan, N. M. A novel focal tversky loss function with improved attention u-net for lesion segmentation. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* 683–687. <https://doi.org/10.1109/ISBI.2019.8759329> (2019).
35. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2261–2269. <https://doi.org/10.1109/CVPR.2017.243> (2017).
36. Tran, D., Bourdev, L., Fergus, R., Torresani, L. & Paluri, M. Learning spatiotemporal features with 3d convolutional networks. In *2015 IEEE International Conference on Computer Vision (ICCV)* 4489–4497 (IEEE Computer Society, Los Alamitos, 2015). <https://doi.org/10.1109/ICCV.2015.510>.
37. Anthony, L. F. W., Kanding, B. & Selvan, R. Carbontracker: Tracking and predicting the carbon footprint of training deep learning models. In *ICML Workshop on Challenges in Deploying and Monitoring Machine Learning Systems*. [arXiv:2007.03051](https://arxiv.org/abs/2007.03051) (2020).
38. Graham-Brown, M. P. M. *et al.* The reproducibility of cardiac magnetic resonance imaging measures of aortic stiffness and their relationship to cardiac structure in prevalent haemodialysis patients. *Clin. Kidney J.* **11**, 864–873 (2018).
39. Bai, W. *et al.* A population-based phenome-wide association study of cardiac and aortic structure and function. *Nat. Med.* **26**, 1654–1662 (2020).
40. Francis, C. M. *et al.* Genome-wide associations of aortic distensibility suggest causality for aortic aneurysms and brain white matter hyperintensities. *medRxiv*. <https://www.medrxiv.org/content/early/2021/09/03/2021.09.01.21262202.full.pdf> (2021).
41. Amodei, D. & Hernandez, D. AI and compute (2018).
42. Strubell, E., Ganesh, A. & McCallum, A. Energy and policy considerations for deep learning in NLP. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* 3645–3650 (Association for Computational Linguistics, Florence, 2019). <https://doi.org/10.18653/v1/P19-1355>.
43. Lacoste, A., Luccioni, A., Schmidt, V. & Dandres, T. Quantifying the carbon emissions of machine learning. *CoRR*. [arXiv:1910.09700](https://arxiv.org/abs/1910.09700) (2019).
44. Henderson, P. *et al.* Towards the systematic reporting of the energy and carbon footprints of machine learning. *J. Mach. Learn. Res.* **21**, 1–43 (2020).
45. Patterson, D. A. *et al.* Carbon emissions and large neural network training. *CoRR*. [arXiv:2104.10350](https://arxiv.org/abs/2104.10350) (2021).
46. Siouras, A. *et al.* Knee injury detection using deep learning on MRI studies: A systematic review. *Diagnostics* **12**, 537 (2022).
47. Finlayson, S. G. *et al.* Adversarial attacks on medical machine learning. *Science* **363**, 1287–1289 (2019).
48. Zhang, H., Gao, Z., Zhang, D., Hau, W. K. & Zhang, H. Progressive perception learning for main coronary segmentation in x-ray angiography. *IEEE Trans. Med. Imaging* **42**, 864–879. <https://doi.org/10.1109/TMI.2022.3219126> (2023).

49. Liu, X. *et al.* Motion estimation based on projective information disentanglement for 3d reconstruction of rotational coronary angiography. *Comput. Biol. Med.* **157**, 106743. <https://doi.org/10.1016/j.compbiomed.2023.106743> (2023).
50. Al-Hussaini, A. *et al.* Chronic infarct size after spontaneous coronary artery dissection: Implications for pathophysiology and clinical management. *Eur. Heart J.* **41**, 2197–2205 (2020).
51. Webb, D. R. *et al.* A randomized, open-label, active comparator trial assessing the effects of 26 weeks of liraglutide or sitagliptin on cardiovascular function in young obese adults with type 2 diabetes. *Diabetes Obes. Metab.* **22**, 1187–1196 (2020).
52. Gulsin, G. S. *et al.* Effects of Low-Energy diet or exercise on cardiovascular function in Working-Age adults with type 2 diabetes: A prospective, randomized, Open-Label, blinded end point trial. *Diabetes Care* **43**, 1300–1310 (2020).
53. Xinapse Systems Ltd. Jim 9 software.
54. Agarap, A. F. Deep learning using rectified linear units (relu). [arXiv:1803.08375](https://arxiv.org/abs/1803.08375) (2018).
55. Ioffe, S. & Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. [arXiv:1502.03167](https://arxiv.org/abs/1502.03167) (2015).
56. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
57. Kingma, D. P. & Ba, J. A method for stochastic optimization. [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2017).
58. He, K., Zhang, X., Ren, S. & Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. [arXiv:1502.01852](https://arxiv.org/abs/1502.01852) (2015).
59. Lin, T.-Y., Goyal, P., Girshick, R., He, K. & Dollár, P. Focal loss for dense object detection. In *2017 IEEE International Conference on Computer Vision (ICCV)* 2999–3007. <https://doi.org/10.1109/ICCV.2017.324> (2017).
60. Salehi, S. S. M., Erdogmus, D. & Gholipour, A. Tversky loss function for image segmentation using 3d fully convolutional deep networks. In *Machine Learning in Medical Imaging* (eds Wang, Q., Shi, Y., Suk, H.-I. & Suzuki, K.) 379–387 (Springer International Publishing, Cham, 2017).
61. Bland, J. M. & Altman, D. G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1**, 307–310 (1986).
62. Average CO₂ emissions from newly registered motor vehicles in Europe—ea.europa.eu. <https://www.eea.europa.eu/data-and-maps/indicators/average-co2-emissions-from-motor-vehicles/assessment-1>. Accessed 2 Aug 2023.

Author contributions

T.A.B.: Writing - original draft, Investigation, Methodology, Software, Visualisation. K.S.P., M.P.M.G.-B., M.M., A.S., J.W., D.A. and D.G.: Data curation, Writing—review & editing. M.J.D., B.W. and M.B.: Resources, Writing—review & editing, G.P.M.: Supervision, Writing—review & editing, Resources. A.G.: Funding acquisition, Supervision, Conceptualisation, Methodology, Investigation, Writing—review & editing, Resources.

Funding

Tuan Aqeel Bohoran is funded by the European Union's Horizon 2020 reasearch and innovation programme under the Marie Skłodowska-Curie Grant Agreement No 801604.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023